2007

A potential bias in the temporal method for estimating Ne in admixed populations under natural selection

Hitoshi Araki
Oregon State University Corvallis, arakih@science.oregonstate.edu

Robin Waples
NOAA, robin.waples@noaa.gov

Michael S. Blouin
Oregon State University Corvallis, blouinm@science.oregonstate.edu

Molecular Ecology (2007) **16**, 2261–2271 doi: 10.1111/j.1365-294X.2007.03307.x

A potential bias in the temporal method for estimating N_e in admixed populations under natural selection

HITOSHI ARAKI,* ROBIN S. WAPLES† and MICHAEL S. BLOUIN*
Department of Zoology, 3029 Cordley Hall, Oregon State University Corvallis, OR 97331, USA, †Northwest Fisheries Science Center, 2725 Montlake Blvd. East, Seattle, WA 98112, USA

Abstract

Indirect genetic methods are frequently used to estimate the effective population size (N_e) or effective number of breeders (N_b) in natural populations. Although assumptions behind these methods are often violated, there have been few attempts to evaluate how accurate these estimates really are in practice. Here we investigate the influence of natural selection following a population admixture on the temporal method for estimating N_e. Our analytical and simulation results suggest that N_e is often underestimated in this method when subpopulations differ substantially in allele frequencies and in reproductive success. The underestimation is exacerbated when true N_e and the fraction of the low-fitness group are large. As an empirical example, we compared N_b estimated in natural populations of steelhead trout (*Oncorhynchus mykiss*) using the temporal method ($\hat{N}_{b[temp]}$) with estimates based on direct demographic methods ($\hat{N}_{b[demo]}$) and the linkage disequilibrium method ($\hat{N}_{b[LD]}$). While $\hat{N}_{b[LD]}$ was generally in close agreement with $\hat{N}_{b[demo]}$, $\hat{N}_{b[temp]}$ was much lower in sample sets that were dominated by nonlocal hatchery fish with low reproductive success, as predicted by the analytical results. This bias in the temporal method, which arises when genes associated with a particular group of parents are selected against in the offspring sample, has not been widely appreciated before. Such situations may be particularly common when artificial propagation or translocations are used for conservation. The linkage disequilibrium method, which requires data from only one sample, is robust to this type of bias, although it can be affected by other factors.

Keywords: effective population size, hatchery program, migration, natural selection, steelhead, temporal method

Received 22 October 2006; revision received 10 January 2007; accepted 29 January 2007

Introduction

The effective population size per generation (N_e) and the effective number of breeders per year (N_b) influence important processes such as genetic drift, inbreeding, and response to selection in natural populations (Hedrick 2005). Although these parameters are of central importance to evolutionary and conservation biology, obtaining reliable estimates of these parameters is elusive at best. Direct calculation of estimates of effective population size ($\hat{N}_e$) is often impossible in natural populations because of the lack of demographic information (e.g. sex ratio, fluctuation of N, and mean and variance of family size per parent, Crow & Kimura 1970). In the absence of extensive pedigree data for estimating these parameters, researchers often rely on indirect genetic methods for estimating N_e. The temporal method is the most commonly used genetic method for estimating effective size. This method estimates variance effective population size based on allele frequency change between samples taken at two or more points in time, and it assumes that observed changes in allele frequency result only from random genetic drift and sampling error (Krimbas & Tsakas 1971; Nei & Tajima 1981; Waples 1989). The linkage disequilibrium (LD) method calculates inbreeding effective population size based on the extent of LD between pairs of loci (Hill 1981) and, unlike the temporal method, can be (but rarely has been) used when only a single sample from a population is available (Bartley *et al*. 1992; Bucci *et al*. 1997).

Correspondence: Hitoshi Araki, Fax: 541-737-0501; E-mail: arakih@science.oregonstate.edu

In natural populations, the assumptions behind these methods are often violated. For example, these genetic methods assume a panmictic population. However, some level of structure is common in natural populations (e.g. Campbell *et al*. 2006; Piggott *et al*. 2006), and migration among subpopulations can bias the temporal N_e estimates substantially (Wang & Whitlock 2003). Selective neutrality at the marker loci is another important assumption. Even if the markers themselves are selectively neutral, natural selection on sites linked to the neutral marker loci can affect N_e at the neutral loci (Charlesworth *et al*. 1995; Santiago & Caballero 1998). Thus, if some of the marker loci are directly or indirectly affected by natural selection, true N_e at each locus may vary (Goldringer & Bataillon 2004).

Natural selection might affect not only the true N_e at each locus but also the estimation of N_e in some cases. One such case (and one which has not been documented previously) is natural selection following a population admixture. When two or more groups of breeders differ in both allele frequency and mean reproductive success, differences in allele frequency between parents and offspring result from both random genetic drift and differential reproductive success, with the result that the temporal method underestimates N_e. An example would be a population that receives an influx of immigrants that have low reproductive success in the new environment. Local adaptation seems to be the rule, rather than the exception, in natural populations (Mousseau *et al*. 2000), so that situation probably arises frequently. That situation also arises when natural populations are deliberately supplemented with individuals raised in an artificial environment (e.g. Fleming & Peterson 2001; Joron & Brakefield 2003; McGinnity *et al*. 2003). The individuals that are artificially introduced often suffer from low fitness in the wild via local adaptation and/or domestication selection, and the strength of natural selection against the supplemented individuals can be substantial (e.g. McGinnity *et al*. 2003; Araki *et al*. 2006). N_e in supplemented populations is of special interest in conservation biology, and theoretical predictions for effects of artificial supplementation on N_e have extensively been addressed (Ryman & Laikre 1991; Waples & Do 1994; Wang & Ryman 2001; Duchesne & Bernatchez 2002). However, the effects of natural selection following a population admixture on the genetic estimation of N_e have not been examined, and there have been few empirical attempts to compare $\hat{N}_e$ based on different methods in actual, supplemented populations (Ardren & Kapuscinski 2003; Miller & Waits 2003).

In this study, we use analytical and empirical approaches and computer simulations to assess the consequences of using genetic methods to estimate effective size in a population that consists of genetically differentiated groups under natural selection. First, we analytically explore the conditions under which the temporal method will give biased estimates assuming a one-generation (parent–offspring) data set. Next, we use Monte Carlo simulations to quantitatively evaluate bias in multigeneration data sets with and without recurrent migration. Finally, we compare the estimates of effective number of breeders per year ($\hat{N}_b$) using the temporal and LD methods with $\hat{N}_b$ based on actual demographic data, in two populations of steelhead trout (*Oncorhynchus mykiss*) in the Hood River. The demographic estimates of N_b in these populations have been obtained from pedigree data in a previous study (Araki *et al*. 2007). The genetic data from these populations, composed of parental samples and their offspring cohorts, allows us direct comparisons among $\hat{N}_b$ using different methods. Our results demonstrate the potential for a substantial downward bias in the temporal estimates of effective population size under certain conditions.

Methods

Analytical model and sampling schemes considered

In the analytical approach, we consider a discrete generation model in which a parental generation is distinguishable from the progeny generation. For simplicity, we focus on the moment-based temporal method (Waples 1989) in this study. MATHEMATICA version 4.1 (Wolfram) was used to calculate analytical results.

The temporal method

The temporal method was applied to each pair of parent and offspring samples to obtain the N_b estimate ($\hat{N}_{b[temp]}$). Because each offspring sample represents a single cohort that was produced by a single parental run year, we could treat the samples as parent–offspring samples from a population with discrete generations. This sample structure simplifies the interpretation of results. Under this sampling regime, estimated N_b always represents the effective number of spawners in the parental year (Waples 2005). Our nonlethal sampling method in the empirical data from steelhead allowed the sampled individuals to reproduce and thus conforms to Plan I sampling in Waples (1989). N_b was estimated by the standardized variance in allele frequency change ($\hat{F}$, Nei & Tajima 1981) as

$$\hat{N}_{b[temp]} = \frac{1}{2[\hat{F} - 1/(2S_P) - 1/(2S_O) + 1/\hat{N}]}, \qquad \text{(eqn 1)}$$

$$\hat{F} = \frac{1}{L}\sum_{i=1}^{L}\frac{(q_{P,i} - q_{O,i})^2}{(q_{P,i} + q_{O,i})/2 - q_{P,i}q_{O,i}}, \qquad \text{(eqn 2)}$$

where S_P and S_O are sample sizes of parents and offspring, and $q_{P,i}$ and $q_{O,i}$ are frequencies of allele i ($i = 1, 2, \ldots L$) in the parent and offspring samples, respectively. $\hat{N}$ represents

Table 1 Genetic and demographic estimates of effective numbers of breeders in two steelhead populations

				Temporal estimation		LD estimation	
Run year	$\hat{N}$*	$\hat{N}_{b[demo]}$*	%Hat*	$\hat{N}_{b[temp]}$ (95% CI)	$\hat{N}_{b[temp]}/\hat{N}_{b[demo]}$	$\hat{N}_{b[LD]}$ (95% CI)	$\hat{N}_{b[LD]}/\hat{N}_{b[demo]}$
Summer run							
Su94	2506	547	89.7**	33 (23–44)	0.06	210 (201–219)	0.38
Su95	1518	204	79.7**	38 (27–51)	0.18	224 (215–233)	1.10
Su96	1902	478	87.8**	32 (23–42)	0.07	423 (405–441)	0.88
Su97	1500	278	84.3**	25 (18–32)	0.09	333 (319–347)	1.20
Su98	927	231	2.9	264 (116–1269)	1.10	367 (352–382)	1.59
Winter run							
Wi91	1781	492	29.0**	120 (84–170)	0.24	396 (372–400)	0.81
Wi92	710	229	1.2	277 (178–447)	1.21	211 (203–218)	0.92
Wi93	760	314	0.5	325 (182–713)	1.04	278 (268–289)	0.89
Wi94	512	234	2.9	210 (131–354)	0.90	173 (165–178)	0.74
Wi95	700	317	40.1***	194 (147–249)	0.61	249 (237–256)	0.78
Wi96	921	501	49.3***	277 (201–370)	0.55	403 (388–418)	0.80
Wi97	868	452	47.2***	301 (209–430)	0.67	490 (472–508)	1.08
Wi98	1395	542	42.6***	276 (183–425)	0.51	498 (482–520)	0.92

$\hat{N}_{b[temp]}$ and $\hat{N}_{b[LD]}$ represent estimates of the number of breeders in each run year from the temporal method and the LD method, respectively. Values indicated in 95%CI represent confidence intervals (CI) at 95%, calculated by methods in Waples (1989) for $\hat{N}_{b[temp]}$ and in Waples (2006) for $\hat{N}_{b[LD]}$.

*data from Araki *et al*. (2007). $\hat{N}$ represents an estimate of the whole population size, $\hat{N}_{b[demo]}$ represents the demographic estimate of N_b, and %Hat represents the percentage of hatchery-born fish among total number of anadromous fish returned to the river in each run year.

**the percentage of traditional hatchery stocks ($\%H_{trad}$).

***the percentage of supplementation hatchery stocks ($\%H_{supp}$).

[Correction added after online publication 10 May 2007: in the preceding table the last column title, $\hat{N}_{b[temo]}/\hat{N}_{b[demo]}$ was corrected to $\hat{N}_{b[LD]}/\hat{N}_{b[demo]}$].

an estimated census population size, which takes the whole population (including nonanadromous fish) into consideration (Table 1, Araki *et al*. 2007). Cutoff level for a minimum allele frequency was set as $q_{P,i}$ or $q_{O,i} > 0.02$ to minimize a bias pointed out by Waples (1989) and Turner *et al*. (2001). Another estimate of the standardized variance in allele frequency change ($\hat{F}_k$, Pollak 1983) was also used to estimate $\hat{N}_b$, but there were few differences between these estimates and the ones reported here (data not shown). The 95% confidence intervals for $\hat{N}_b$ were calculated using the 2.5% and 97.5% cumulative probabilities of the χ^2 distribution according to the method of Waples (1989).

Monte Carlo simulations for multigeneration effects

We used Monte Carlo simulations for analyses of multiple generation effects. For simplicity, we assumed the haploid Wright-Fisher model with discrete generation. The other assumptions are as follows: (i) we consider two loci, one of which is under selection and the other locus is a neutral marker; (ii) either free or no recombination between the loci; and (iii) local adaptation, in which immigrants have lower reproductive success than local residents because all the immigrants have a deleterious allele at the selected locus with relative fitness R ($R \leq 1$). Multigeneration effects are examined in two scenarios, a single-pulse migration and recurrent migration. First, we assumed a pulse of migration of a low-fitness group at the initial generation ($t = 0$), as we did in the analytical approach. Then we estimated the temporal $\hat{N}_e$ in two sampling schemes, depending on the time of the first temporal sampling. In the first sampling scheme, the temporal $\hat{N}_e$ (= $t/\hat{F}$ for the haploid population) is calculated measuring the change in allele frequency between generation 0 and t. This sampling scheme represents a situation in which temporal samples are separated by multiple generations and only the initial sample includes new immigrants. In the second sampling scheme, we measured allele frequency change between generations t-1 and t. In this sampling scheme, therefore, we sampled parents and their offspring, but those two samples were taken a variable number of generations after the migration event. These simulations were designed to evaluate over how many generations a pulse of migration (at $t = 0$) affects a single-generation temporal estimate (measured between generations t-1 and t).

In the recurrent migration scenario, we introduced a constant proportion of immigrants that have low reproductive success every generation. To reach an equilibrium state before the N_e estimation, 100 generations of the simulations were performed in advance of collecting the first sample. We confirmed that allele frequencies (both at selected and marker loci) are stable after the 100 generations. We

obtained simulation results using the first sampling scheme (between generation 0 and t) in this scenario, because results using the second sampling scheme (between generations t-1 and t) were unchanged among different t and close to the results at $t = 1$ in the first sampling scheme because of the equilibrium state. In all the cases, we calculated the temporal $\hat{N}_e$ for 10 generations in each simulation, and $\hat{F}$ were averaged over 10 000 replications in order to obtain $\hat{N}_e$.

Empirical data from steelhead

Here we used the same data set of the Hood River steelhead as in Araki *et al*. (2007) (Table 1). The Hood River basin is a tributary of the Columbia River in the northwest United States. Since 1991, almost every adult steelhead returning to spawn in this river has had scale and fin-snip samples taken at the Powerdale Dam fish trap. This dam is located 4 river miles from the mouth of the Hood River. Steelhead spawn only above the dam, which is a complete upstream barrier to migration for all salmonids. Thus, we have DNA samples from almost every anadromous steelhead that spawned in this river for 15 years.

This river holds two distinct populations (runs) of steelhead, the summer run and the winter run. The run year is the year in which adult fish begin arriving at the river to spawn, and is the calendar year before they actually spawn. In this study, we investigated N_b in five run-year samples of summer run (run years 1994 to 1998; hereafter, Su94 to Su98) and eight run-year samples of winter run (run years 1991 to 1998; hereafter, Wi91 to Wi98) in order to compare the results with demographic estimates of N_b in Araki *et al*. (2007). Both runs breed in the spring, but in different forks of the river. Only a few hybridization events were identified between populations in the parentage analysis (Araki *et al*. 2007). To identify a single offspring cohort from a specific run year of parents, brood years of the returning adults were examined via scale-reading. The total numbers of individuals used in this study were 4561 for summer-run parents, 2217 for summer-run offspring cohorts, 3945 for winter-run parents, and 5991 for winter-run offspring cohorts. Note that a single offspring cohort is composed of multiple run years of returning adults, because of the variable age at spawning in the steelhead populations (normally 3–7 years old).

Hatchery programs

Traditional hatchery stocks (H_{trad}) are produced using nonlocal brood stocks that have spent multiple generations in hatcheries. Supplementation hatchery stocks (H_{supp}) are produced using local, wild fish as brood stock. Our data sets are composed of four run years that include large proportions of H_{trad} (80–90% in Su94–97) and one run year with a modest proportion of H_{trad} (29% in Wi91), four run years that are almost exclusively wild fish (Su98, Wi92–94, $\%H_{trad} < 3\%$), and four run years in which the proportions of H_{supp} ($\%H_{supp}$) are close to 50% (Wi95–98) (Table 1). H_{trad} fish have substantially lower reproductive success than the other two types when breeding in the wild (Araki *et al*. 2006).

Microsatellite loci

Genotypes at eight microsatellite loci (Omy1001, Omy1011, Omy1191, Omy77, One108, One2, Ssa407, and Str2) in the samples above were determined in the previous studies (Araki *et al*. 2006, 2007). Our data include 97.6% of adult samples that returned to the river between run years 1994 and 2003 (for winter run, 1998–2003 for summer run). These loci are highly polymorphic, and the average expected heterozygosities are about 90% in all the loci. In a previous study (Araki *et al*. 2007), we performed a parentage analysis in order to estimate the mean and variance in the sizes of families produced by the breeders each year, which gives demographic estimates of N_b for each population in each run year. Here we utilized these genetic data to measure the allele frequency changes between parental samples and their offspring cohorts (for the temporal method) and the allelic linkage disequilibrium among the offspring cohorts (for the LD method).

The linkage disequilibrium method

For each offspring cohort, we estimated $N_b(\hat{N}_{b[LD]})$ in the parents from the linkage disequilibrium in their offspring via the LD method (Hill 1981). We used the Burrows method (Weir 1996) to calculate the squared correlation of allele frequencies ($\hat{r}^2$) at pairs of loci. The overall mean $\bar{\hat{r}}^2$ for each sample was computed as the weighted average $\hat{r}^2$ over the $L(L–1)/2$ pairwise comparisons of different gene loci, with L being the number of polymorphic loci. The weights for each locus pair were the relative number of independent alleles used in the comparison. For two loci with K_1 and K_2 alleles, respectively, there are the equivalent of $(K_1 - 1)(K_2 - 1)$ independent allelic comparisons. The total degrees of freedom associated with the overall weighted mean $\bar{\hat{r}}^2$ was computed as $n = \sum(K_i - 1)(K_j - 1)$, with the sum over all pairwise comparisons of loci. To minimize possible bias from low frequency alleles, we only used alleles with a frequency ≥ 0.02. We used the bias correction developed by Waples (2006; random mating model) to estimate effective size from the overall mean $\bar{\hat{r}}^2$. Confidence intervals to $\hat{N}_e$ were computed using equation 12 in Waples (2006) assuming n degrees of freedom. These parametric confidence intervals are undoubtedly too narrow (but by an unknown amount) because the different pairwise comparisons of loci are not independent (Hill 1981; Waples 2006), and the effects of large numbers of alleles per locus on n have not been evaluated.

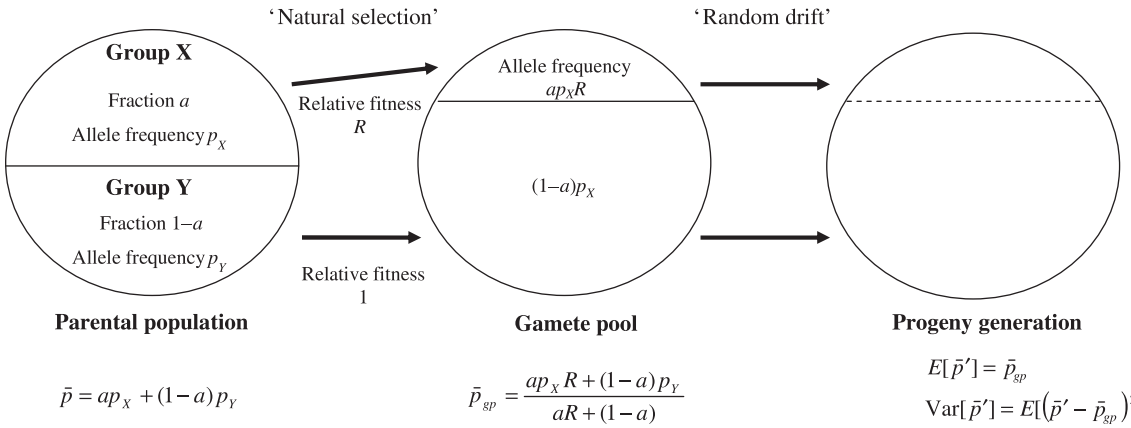

Fig. 1 Schematic diagram of temporally spaced sample sets. In the parental population, fractions of group X and Y are a and $(1 - a)$, and allele frequencies within each group are p_x (in X) and p_y (in Y). Average fitness in group X is R (relative to 1 in group Y). $\bar{p}$, $\bar{p}_{gp}$, $\bar{p}'$ represent average allele frequencies in the parental population, gamete pool, and progeny generation, respectively. $\bar{p}_{gp}$ is deterministically given by natural selection, but $\bar{p}'$ is determined stochastically due to random genetic drift.

The demographic method

The demographic estimates of $N_b(\hat{N}_{b[demo]})$ were obtained using the sex ratio, variance in family size in both sexes in each run year, and the whole (census) population sizes of each run year, as estimated in the previous study (Araki *et al*. 2007). Obviously, the demographic $\hat{N}_b$ is subject to sampling errors and other uncertainties (e.g. parentage assignment errors), and should not be considered as the true N_b. However, the almost complete sampling of anadromous adults in the study and the fact that the demographic method does not require as many assumptions as the genetic methods do (above) suggest that the demographic estimate should be the most reliable among the three estimators in our study.

Comparing estimates of N_b

In comparing estimates of effective size, it is important to ensure that the same time periods are being considered. In the case of parent–offspring samples, the temporal method estimates N_b in the parental generation, and LD measured in the offspring also produces an estimate of N_b in the parental generation (Waples 2005). Thus, the temporal estimates using genetic data from the adult samples spawning the same run year and their offspring cohorts are directly comparable to the LD estimates using the same genetic data from the offspring cohorts, despite the fact that the former applies to the variance effective size and the latter applies to the inbreeding effective size. The demographic method produced an estimate of inbreeding N_b that is also applicable to the parental generation (Araki *et al*. 2007). Therefore, the temporal, LD, and demographic methods all estimate N_b in the adults spawning each run year and can be compared directly.

We compared the genetic estimates of N_b with $\hat{N}_{b[demo]}$ by taking ratios ($\hat{N}_{b[temp]}/\hat{N}_{b[demo]}$ or $\hat{N}_{b[LD]}/\hat{N}_{b[demo]}$). The overall ratio for each population (among run years) was obtained using geometric mean, because the geometric mean provides an unbiased summary statistics (i.e. the geometric mean results in the expected ratio equals one when the expectation of denominator and that of numerator are the same).

Results

Bias in the temporal method using a single-generation data set

We want to evaluate the bias in F (and hence $\hat{N}_e$) that arises when allele frequency differences between parents and offspring are influenced not only by genetic drift but also by differential reproductive success of two groups of parents. For simplicity, we ignore sampling and assume that the entire population is assayed genetically ($\hat{N} = S_P = S_O$). First, we consider one allele in a hypothetical population that is composed of two groups (X and Y) (Fig. 1). If the allele frequency in groups X and Y is p_X and p_Y, respectively, average allele frequency in the parental population ($\bar{p}$) is

$$\bar{p} = ap_X + (1 - a)p_Y, \qquad \text{(eqn 3)}$$

where a is a fraction of group X in the population. Now, let the two groups contribute genes to a large gamete pool that will form the next generation, and let R be the per-capita relative reproductive contribution of group X (immigrants) compared to group Y (local residents). Under these conditions, the average allele frequency in the gamete pool produced by the parental population ($\bar{p}_{gp}$) is

$$\bar{p}_{gp} = \frac{ap_X R + (1-a)p_Y}{aR + (1-a)}, \qquad \text{(eqn 4)}$$

If the next generation is produced by random draw from the gamete pool, the allele frequency in the progeny, $\bar{p}'$, is subject to random genetic drift. We evaluate $(\Delta\bar{p})^2$, where $\Delta\bar{p}$ is the difference in allele frequency between parents and offspring (i.e. over a single-generation and $\Delta\bar{p} = \bar{p}' - \bar{p}$). We note that $\Delta\bar{p}$ can be written as $\Delta\bar{p} = [(\bar{p}' - \bar{p}_{gp}) + (\bar{p}_{gp} - \bar{p})]$, which shows that the allele frequency change has two components, the first related to genetic drift, and the second related to the selective differences between the groups (Fig. 1). For any given initial conditions, the change due to selection ($K = \bar{p}_{gp} - \bar{p}$) is a constant determined by equations 3 and 4, so we can write $\Delta\bar{p} = (\bar{p}' - \bar{p}_{gp}) + K$.

In the diploid model, taking the expectation of $(\Delta\bar{p})^2$ gives us

$$E[(\Delta\bar{p})^2] = E[(\bar{p}' - \bar{p}_{gp})^2] + K^2 = \frac{\bar{p}_{gp}(1-\bar{p}_{gp})}{2N_e} + K^2, \qquad \text{(eqn 5)}$$

where N_e is the effective size in the progeny generation. Standardizing to initial allele frequency (Waples 1989) leads to

$$E\left[\frac{\Delta\bar{p}^2}{\bar{p}(1-\bar{p})}\right] = E[F] = \frac{\bar{p}_{gp}(1-\bar{p}_{gp})}{\bar{p}(1-\bar{p})2N_e} + \frac{K^2}{\bar{p}(1-\bar{p})} \approx \frac{1}{2N_e} + \frac{K^2}{\bar{p}(1-\bar{p})}. \qquad \text{(eqn 6)}$$

This approximation is obtained when $\bar{p}_{gp}(1-\bar{p}_{gp})/\bar{p}(1-\bar{p}) \approx 1$, which is reasonable if K is not too large. Recalling that $K = (\bar{p}_{gp} - \bar{p})$, we see that the second term in equation 6 is essentially F_{ST} between the parental population and the gamete pool it produces. Equation 6 shows that under the scenario considered, the temporal F is biased upwards by approximately the F_{ST} value generated by selective differences between parental groups. This bias is zero if $K = 0$, which occurs if any of the following is true: a is 0 or 1 (only one group); $p_X = p_Y$ (no allele frequency differences between the groups); or $R = 1$ (no selection). Otherwise, selection will cause an allele frequency shift between the parental population and the gamete pool that will upwardly bias F and downwardly bias $\hat{N}_e$. Since K is a constant for given initial conditions, the relative importance of the bias will be greatest when N_e is large and hence the first term small.

Now let

$$Q = K^2/\bar{p}(1-\bar{p}) = \frac{\bar{p}_{gp}^2 + \bar{p}^2 - 2\bar{p}_{gp}\bar{p}}{\bar{p}(1-\bar{p})},$$

and note that $\hat{N}_e = 1/(2F)$ in the diploid model if we ignore sampling in the temporal method ($\hat{N} = S_P = S_O$ in equation 1). Substituting the expected value of F from equation 6 leads to

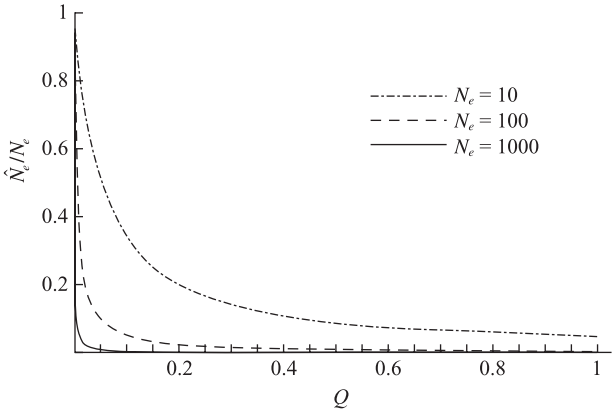

Fig. 2 Effects of Q (= $K^2/\bar{p}(1-\bar{p})$) on $\hat{N}_e/N_e$ in the temporal method. True N_e used here are 10, 100, and 1000 (lines top to bottom). Note that $Q = 0$ when $K = (\bar{p}_{gp} - \bar{p}) = 0$, whereas $Q = 1$ when $\bar{p} = 0.5$ and $\bar{p}_{gp} = 0$ or 1. Curves were generated analytically using equation 8.

$$\hat{N}_e \approx \frac{1}{\frac{1}{N_e} + 2Q} = \frac{N_e}{1 + 2QN_e} \qquad \text{(eqn 7)}$$

and the ratio of $\hat{N}_e$ to true N_e is

$$\frac{\hat{N}_e}{N_e} \approx \frac{1}{1 + 2QN_e}. \qquad \text{(eqn 8)}$$

Equation 8 shows that the estimate of effective size will be unbiased if $Q = 0$, which implies $K = 0$ and which occurs under the conditions described above. As Q increases, however, $\hat{N}_e/N_e$ decreases very quickly, and the bias is stronger when true N_e is large (Fig. 2).

To evaluate the influence of each parameter on the bias, $\hat{N}_e/N_e$ ratios were plotted in various parameter sets (Fig. 3 for examples of $a = 0.5$, $p_x = 0.5$, $p_y = \{0.45, 0.40, 0.30\}$, $R = 0.3$, $N_e = 100$ unless mentioned). Results are as follows: first, natural selection is a major determinant of the level of bias, and the bias is strong when the difference in allele frequency between groups is large as expected (Fig. 3a). When natural selection is very strong (e.g. $R \leq 0.3$), even a slight difference in allele frequency between groups results in a strong bias (top line in Fig. 3a). Second, true N_e affects $\hat{N}_e/N_e$ even more strongly than does natural selection (Fig. 3b). $\hat{N}_e/N_e$ declines very quickly as N_e increases, indicating that $\hat{N}_e$ does a poor job of tracking true N_e. This figure illustrates the effect of N_e seen in equation 7, in which $\hat{N}_e$ approaches $1/2Q$ as N_e increases. Third, the fraction of lower-fitness group (a) maximizes the bias when a is intermediate, but this effect is relatively minor when the difference in allele frequency between groups is small. When the difference in allele frequency between groups is large, however, the bias is stronger if a is close to one rather than close to zero (Fig. 3c). It means that $\hat{N}_e$ is more strongly biased when larger fraction of parents have low reproductive success.

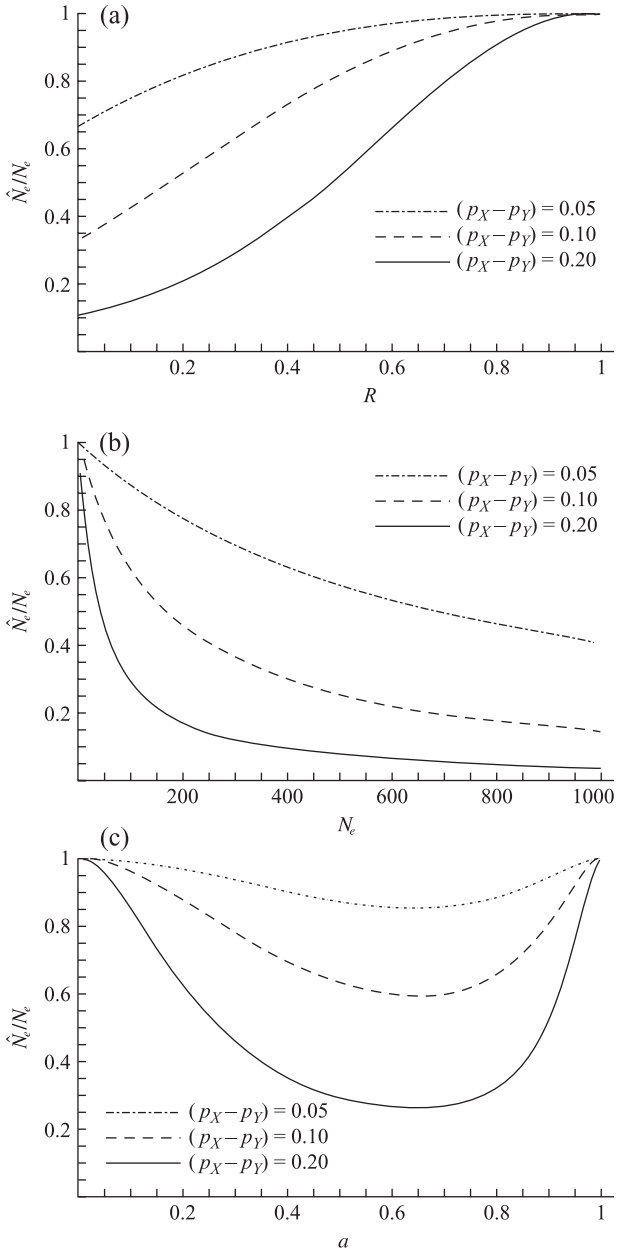

Fig. 3 Effects of various parameters on $\hat{N}_e/N_e$ in the temporal method. (a) Effect of natural selection parameter (R) on $\hat{N}_e/N_e$. The other parameters are: $a = 0.5$, $p_x = 0.5$, $p_y = \{0.45, 0.40, 0.30\}$ (lines top to bottom), $N_e = 100$. (b) Effect of true effective population size (N_e) on $\hat{N}_e/N_e$. The other parameters are: $a = 0.5$, $p_x = 0.5$, $p_y = \{0.45, 0.40, 0.30\}$ (lines top to bottom), $R = 0.3$. (c) Effect of fraction of group X (a) in the parental population on $\hat{N}_e/N_e$. The other parameters are: $p_x = 0.5$, $p_y = \{0.45, 0.40, 0.30\}$ (lines top to bottom), $R = 0.3$, $N_e = 100$. $\hat{N}_e/N_e = 1$ when there is no bias. Curves were generated analytically using equations 3, 4, and 8.

Multigeneration effects

All the conclusions above were based on single-generation data (parent–offspring samples), in which only the parent sample includes new immigrants. In the simulations of a haploid model, we examined cases of $a = 0.5$, $p_x = 0.5$, $p_y = 0.30$, $R = \{0.3, 0.5, 0.8\}$ and $N_e = \{200, 2000\}$ to compare results with those using the analytical approach using the diploid model above. In the single-pulse migration scenario at $t = 0$, we further considered two sampling schemes, sampling at generation 0 and t, and sampling at generation t-1 and t (Methods).

Pulse migration. In the first sampling scheme (samples taken in generations 0 and t), simulation results showed that the bias remained large even when the marker was not linked to the selected locus (Fig. 4a). When there was no recombination between the loci (Fig. 4b), $\hat{N}_e/N_e$ even decreased as t increased in some cases. These results presumably reflect the fact that the change in allele frequency due to natural selection accumulates through the generations, until either recombination sets the marker locus free from the selection or the deleterious allele is eliminated from the population.

In the second sampling scheme (samples taken in generations t-1 and t), the bias disappeared quickly (within a few generations) when the marker locus was not linked to the locus under selection (Fig. 4c). When there was no recombination (Fig. 4d), however, the bias remained long and recovery from the bias was slower under weak selection (i.e. when R is large).

Recurrent migration. In the first sampling scheme with recurrent migration, $\hat{N}_e/N_e$ increased linearly as t increased (Fig. 4e,f). $\hat{N}_e/N_e$ was generally larger than one when t was large. The overestimation of N_e under this scenario is consistent with results obtained by Wang & Whitlock (2003), which assumed no selection (see Discussion). In the second sampling scheme, $\hat{N}_e/N_e$ values were independent of t and similar to those for $t = 1$ in Fig. 4e (free recombination) and Fig. 4f (no recombination), because the population was at the equilibrium state before the sampling (Methods). In general, the downward bias was smaller than occurred under the pulse migration, as expected from the fact that recurrent migration with constant allele frequency reduces the allele frequency differences between immigrants and local residents. When selection was very strong (e.g. $R = 0.3$) and when there was no recombination, however, the bias stayed large (Fig. 4e,f).

$\hat{N}_b$ *in steelhead populations*

In the natural populations of steelhead, $\hat{N}_{b[temp]}$ and $\hat{N}_{b[LD]}$ were evaluated by comparing them with estimates of N_b from the demographic method ($\hat{N}_{b[demo]}$, see Methods). The temporal method generally underestimated N_b relative to estimates from the demographic method (Table 1). The downward bias in $\hat{N}_{b[temp]}$ was very strong in Su94–97 and Wi91 but less so in Wi95–98. Estimates of $\hat{N}_{b[temp]}$ in Su98

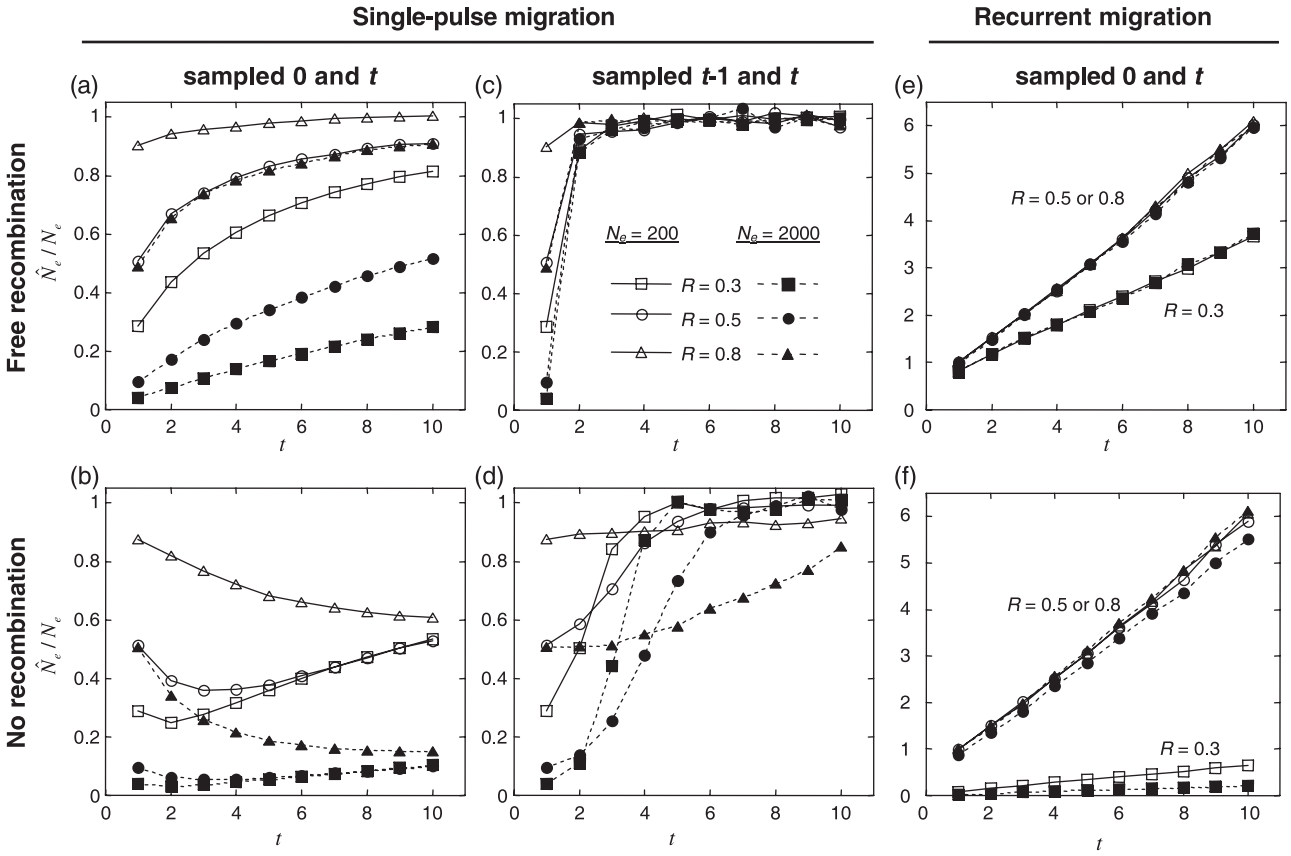

Fig. 4 Simulation results for multigeneration effects on temporal estimates ($\hat{N}_e$) in admixed populations subject to selection. Monte Carlo simulations on a haploid model were used. Upper panels: free recombination between loci; lower panels: no recombination. (a–d): a single-pulse migration at generation 0. (a,b) temporal sampling at generations 0 and t. In these cases, the downward bias decreases only slowly as t increases, and the bias actually increases when there is no recombination and selection is weak. (c,d) temporal sampling at generations t-1 and t. In these cases, the downward bias disappears quickly in general, but remains long when there is no recombination and selection is weak. (e,f): recurrent migration every generation, with 100 generations of migration before collecting data to ensure equilibrium conditions; temporal sampling at generations 0 and t. In these cases, $\hat{N}_e/N_e$ increases linearly with t, and the bias is generally upward except when there is no recombination and selection is strong. Other parameters are $a = 0.5$, $p_x = 0.5$, and $p_y = 0.30$, which correspond to Fig. 3. In all the cases, the results represent $\hat{N}_e/N_e$ calculated over 10 000 replications.

and Wi92–94 were similar to the demographic estimates. These sets of run years perfectly match parental samples with large proportions of H_{trad} (Su94–97 and Wi91), those with about 50% of H_{supp} (Wi95–98), and those with almost no hatchery fish (Su98 and Wi92–94) (Table 1). The ratios of $\hat{N}_{b[temp]}$ to $\hat{N}_{b[demo]}$ were especially low (< 10%) for Su94, Su96, and Su97, in which $\%H_{trad}$ were the largest (> 80%).

Estimates of N_b from the LD method were much closer to, and were not consistently higher or lower than, the demographic estimates (74–159% of demographic estimates; the one exception being 38% in Su94). Seven out of eight sample sets in the winter run showed $\hat{N}_{b[LD]} < \hat{N}_{b[demo]}$, but the differences were relatively minor (the geometric mean of the ratio $\hat{N}_{b[LD]}/\hat{N}_{b[demo]}$ in winter run = 0.85), and $\hat{N}_{b[LD]}$ was not consistently smaller than $\hat{N}_{b[demo]}$ in summer run (the geometric mean = 0.93 in summer run, and = 1.16 when Su94 is excluded). Su94 was notable in having the largest run size of anadromous parents, the highest proportion of hatchery breeders, and the fewest returning offspring per capita (Araki *et al*. 2007), but it is not immediately clear why these factors should lead to an unusually low $\hat{N}_{b[LD]}/\hat{N}_{b[demo]}$ ratio.

Discussion

The analytical and simulation results suggest that the temporal method for estimating N_e will be strongly biased under many circumstances of admixed populations under natural selection. Although our results used the moment-based temporal method, qualitatively similar biases would be expected for more recent likelihood-based implementations of the temporal method (e.g. Wang 2001; Berthier *et al*. 2002), since the latter methods also depend on the magnitude of allele frequency change over time.

The bias is particularly strong when N_e is large because the absolute magnitude of bias in F (Q in equation 8) is relatively more important in large populations (Fig. 2, Fig. 3b, and Fig. 4). The insensitivity of $\hat{N}_e$ to N_e (small $\hat{N}_e$ regardless of the true N_e) could lead to inappropriate conclusions, not only about the degree of genetic drift and effectiveness of natural selection in natural populations, but also about the conservation status of the populations.

The Monte Carlo simulations demonstrate that the level of bias in the temporal $\hat{N}_e$ further depends on sampling schemes, recombination rates around marker loci, and the history of migration. Our data suggest that when the first temporal sample contains low-fitness migrants from a pulse migration, the bias does not disappear even though the second temporal sample is collected several generations later (Fig. 4a). In fact, the bias actually increases over time (as the second temporal sample is collected in later generations), if marker loci are linked to sites under selection (Fig. 4b). On the other hand, the bias is negligible when the pulse of migration occurs several generations earlier than the first temporal sample is collected and the marker loci are not linked to the sites under selection (Fig. 4c). It is interesting that if the marker loci are linked to sites under selection, recovery from the bias is slower when the selection is rather weaker (Fig. 4d). We believe that this delay in recovery from the bias is because the population takes longer to eliminate the deleterious allele introduced by the migrants when selection is less effective. Indeed, the average generation at which the deleterious allele frequencies first became < 0.1 was $t = 2$ for $R = 0.3$, $t = 4$ for $R = 0.5$, and $t = 10$ for $R = 0.8$ when N_e was 2000 in the condition of Fig. 4d.

Ultimately, the source of the bias described in this study arises because of the violations of two explicit assumptions of the temporal method — that the target population is closed to migration, and that the marker loci behave neutrally. In the absence of selection, migration can have different effects depending on its timing relative to the samples. If immigration begins after the first sampling period, then the final allele frequencies can be changed beyond that expected from the local N_e, causing an underestimate of N_e. On the other hand, if the target population receives a constant rate of immigration from a large stable source population, allele frequency changes will be underestimated as sampling interval increases, which results in an overestimate of N_e (Fig. 4e,f). This latter effect occurs because genes at constant frequency arriving from the source population represent a form of genetic inertia to the target population, and the local gene frequencies do not drift as much as would be predicted from the local N_e (Wang & Whitlock 2003). In our simulation, the linear increase of $\hat{N}_e/N_e$ was observed even in a case of no selection ($R = 1$, data not shown), suggesting that the overestimation is indeed caused by the recurrent migration itself.

Here we have shown that when immigration begins before the first sampling period and is coupled with selection, the general result is a downward bias in estimates of N_e. Even under Wang & Whitlock's (2003) scenario (recurrent migration from a constant source), selection can still cause an underestimation of N_e under certain conditions (e.g. when selection is very strong and the marker is linked to a selected site; Fig. 4f).

Under recurrent migration, the net direction of bias probably depends on the migration–selection balance, which affects the level of genetic differentiation between migration-source and local populations. Indeed, in all the cases in which $\hat{N}_e/N_e$ is close to one at $t = 1$ in Fig. 4(e,f), allele frequencies at the marker locus in the local population (p_y) were very close to those in the migrants (p_x) from the beginning of the temporal N_e estimations, and the deleterious allele at the selected locus was nearly fixed in the populations due to massive recurrent migration we assumed ($a = 0.5$). These simulation results are consistent with expectations from the analytical results (equation 8), which shows that the magnitude of bias is inversely related to $Q = K^2/\bar{p}(1-\bar{p}) \approx F_{ST}$ between the parental population and the gamete pool it produces. In the pulse migration scenario, Q is very large at $t = 1$ (because of the fixed allelic difference between immigrants and local residents) and remains relatively large for any pair of samples taken multiple generations apart as long as the first sample includes immigrants with divergent allele frequencies. When the samples are taken only one generation apart and the first sample is taken some time after the pulse migration, Q is smaller because the maladapted migrant alleles have been reduced in frequency before sampling begins. In the recurrent migration scenario, Q is always small because at equilibrium the frequency of the maladapted allele in the local population approaches that in the immigrants.

In the Monte Carlo simulations, we assumed a simple haploid model, which we believe is also relevant to a diploid model without dominance. If the deleterious allele at the selected locus is recessive, the influence of natural selection should persist longer in the population and the recovery from the bias should be delayed.

In the Hood River steelhead populations, the fact that the strongest deviation of $\hat{N}_{b[temp]}$ was observed in run years with high proportions of H_{trad} fish suggests that $\hat{N}_{b[temp]}$ was biased due to the very poor reproductive success of H_{trad} fish (6–45% of wild-born fish), as previously identified for some of these run years (Su95–96 and Wi91, Araki *et al*. 2006). According to the analytical results, if the parental generation includes a large number of genetically distinct and less-fit (nonlocal) fish, changes in allele frequency between parental and offspring samples will be inflated and $\hat{N}_{b[temp]}$ will be underestimated. That wild-born parents and traditional hatchery-born parents differ detectably in allele frequencies in Hood River steeelhead supports this

hypothesis (F_{ST} ~0.02, unpublished data). Moderate levels of deviation of $\hat{N}_{b[temp]}$ observed in the run years with 50% of supplementation (local) hatchery-born fish indicate that the temporal method is quite sensitive to the different genetic composition of the temporally spaced samples and natural selection.

This type of bias is expected to be largest in years in which the hatchery fish represent a substantial fraction of parents but leave few offspring — as what occurred in Su94 and Su96. The LD method, which is based entirely on data for the sample of offspring that are actually produced, should not be affected by this particular bias. However, the LD method can be affected by other factors, including population mixture and admixture (Nei & Li 1973; Waples & Smouse 1990; Vitalis & Couvet 2001). If the offspring sample includes individuals from two different gene pools, or if the offspring are progeny of matings between genetically differentiated populations, then the magnitude of LD will be inflated by disequilibrium due to population mixture/admixture (a two-locus Wahlund effect, Sinnock 1975) as well as genetic drift. The inflated LD will, in turn, downwardly bias estimates of effective size. This effect would be strongest when the different parental groups are genetically very divergent and have comparable contributions to the offspring generation (Waples & Smouse 1990). This factor might be partially responsible for the slightly reduced $\hat{N}_{b[LD]}$ in some winter run samples (except Wi92–94, which had no hatchery fish) and the low value for Su94. Our results, however, suggest that in the steelhead example any biases associated with $\hat{N}_{b[LD]}$ were generally much smaller than those associated with $\hat{N}_{b[temp]}$.

The Hood River steelhead program is a common example of artificial propagation. In this population, a local population receives episodic infusion of genetically differentiated (nonlocal) immigrants, and the immigrants have very low reproductive success. The maladaptation of the nonlocal immigrants in turn could help maintain the local adaptation and the population genetic differentiation. Under those conditions, allele frequency differences between the parents (locals + immigrants) and their offspring (or subsequent generations) would be larger than those between local parents and their offspring. As a consequence, a temporal estimate of N_e would be downwardly biased, just as what was found in this study. Given the demonstrated importance of local adaptation in a wide range of taxa (Mousseau *et al*. 2000), this potential source of bias in the temporal method might be more common than has previously been appreciated. A slight variant of this scenario is suggested by the widespread evidence for outbreeding depression in F_2 and subsequent generations in hybridized populations of many taxa that do not show a reduction in fitness (and which might even exhibit heterosis) in the F_1 generation (reviewed by Tallmon *et al*. 2004). Under this scenario, exogenous alleles present in immigrants would be selectively removed from the population in the F_2 and subsequent generations, leading to a similar type of downward bias in temporal estimates based on samples of parents (locals and immigrants) and their descendents.

Conclusions

Using both analytical and experimental approaches and computer simulations, we found strong downward biases in N_e estimates from the temporal method when genes associated with a particular group of parents are selected against in the offspring sample. Our results illustrate some caveats that should be kept in mind in using genetic methods to estimate effective size of composite populations that include two or more parental groups with (potentially) differences in mean reproductive success. Although our example involved hatchery and wild populations of fish, the problem is more general and could apply to a wide range of situations of conservation interest. For example, translocation programs are increasingly being used with endangered species (Tallmon *et al*. 2004), and these often would lead to the same type of local/nonlocal sampling issues considered here. Even in entirely natural systems, dynamic processes can lead to episodic infusions of migrant individuals into local populations, where the migrants often have greatly reduced reproductive success. Application of the temporal method to such systems could lead to the same type of bias described here. In contrast, substantial biases in the LD method would be more likely to occur if the sample has been unwittingly taken of progeny from two or more gene pools, or if the sample is from a recently admixed population.

Acknowledgements

We thank M. Ford for useful discussion, E. Olsen, B. Cooper and W. R. Ardren for technical help, and K. Kostow for arranging initial funding for the work. This research was funded by contracts to M.S.B. from the Bonneville Power Administration and the Oregon Department of Fisheries and Wildlife.

References

Araki H, Ardren WR, Olsen E, Cooper B, Blouin MS (2006) Reproductive success of captive-bred steelhead trout in the wild: evaluation of three hatchery programs in the Hood River. *Conservation Biology*, **21**, 181–190.

Araki H, Waples RS, Ardren WR, Cooper B, Blouin MS (2007) Effective population size of steelhead trout: influence of variance in reproductive success, hatchery programs, and genetic compensation between life-history forms. *Molecular Ecology*, **16**, 953–966.

Ardren WR, Kapuscinski AR (2003) Demographic and genetic estimates of effective population size (N_e) reveals genetic compensation in steelhead trout. *Molecular Ecology*, **12**, 35–49.

Bartley D, Bagley M, Gall G, Bentley M (1992) Use of linkage disequilibrium data to estimate effective size of hatchery and natural fish populations. *Conservation Biology*, **6**, 365–375.

Berthier P, Beaumont MA, Cornuet JM, Luikart G (2002) Likelihood-based estimation of the effective population size using temporal changes in allele frequencies: a genealogical approach. *Genetics*, **160**, 741–751.

Bucci G, Vendramin GG, Lelli L, Vicario F (1997) Assessing the genetic divergence of *Pinus leucodermis* Ant. endangered populations: use of molecular markers for conservation purposes. *Theoretical and Applied Genetics*, **95**, 1138–1146.

Campbell P, Schneider CJ, Adnan AM, Zubaid A, Kunz TH (2006) Comparative population structure of Cynopterus fruit bats in peninsular Malaysia and southern Thailand. *Molecular Ecology*, **15**, 29–47.

Charlesworth D, Charlesworth B, Morgan MT (1995) The pattern of neutral molecular variation under the background selection model. *Genetics*, **141**, 1619–1632.

Crow JF, Kimura M (1970) *An Introduction to Population Genetics Theory*. Harper & Row, New York.

Duchesne P, Bernatchez L (2002) An analytical investigation of the dynamics of inbreeding in multi-generation supportive breeding. *Conservation Genetics*, **3**, 45–58.

Fleming IA, Peterson E (2001) The ability of released hatchery salmonids to breed and contribute to the natural productivity of wild populations. *Nordic Journal of Freshwater Research*, **75**, 71–98.

Goldringer I, Bataillon T (2004) On the distribution of temporal variations in allele frequency: consequences for the estimation of effective population size and the detection of loci undergoing selection. *Genetics*, **168**, 563–568.

Hedrick PW (2005) *Genetics of Populations*. Jones & Bartlett, Boston.

Hill WG (1981) Estimation of effective population size from data on linkage disequilibrium. *Genetical Research*, **38**, 209–216.

Joron M, Brakefield PM (2003) Captivity masks inbreeding effects on male mating success in butterflies. *Nature*, **424**, 191–194.

Krimbas CB, Tsakas S (1971) The genetics of *Dacus oleae*. V. Changes of esterase polymorphism in a natural population following insecticide control — selection or drift? *Evolution*, **25**, 454–460.

McGinnity P, Prodohl P, Ferquson A *et al.* (2003) Fitness reduction and potential extinction of wild populations of Atlantic salmon, *Salmo salar*, as a result of interactions with escaped farm salmon. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, **270**, 2443–2450.

Miller CR, Waits LP (2003) The history of effective population size and genetic diversity in the Yellowstone grizzly (*Ursus arctos*): implications for conservation. *Proceedings of National Academy of Sciences, USA*, **100**, 4334–4339.

Mousseau TA, Sinervo B, Endler JA (2000) *Adaptive Genetic Variation in the Wild*. Oxford University Press, New York.

Nei M, Li WH (1973) Linkage disequilibrium in subdivided populations. *Genetics*, **75**, 213–219.

Nei M, Tajima F (1981) Genetic drift and estimation of effective population size. *Genetics*, **98**, 625–640.

Piggott MP, Banks SC, Taylor AC (2006) Population structure of brush-tailed rock-wallaby (*Petrogale penicillata*) colonies inferred from analysis of faecal DNA. *Molecular Ecology*, **15**, 93–105.

Pollak E (1983) A new method for estimating the population size from allele frequency changes. *Genetics*, **104**, 531–548.

Ryman N, Laikre L (1991) Effects of supportive breeding on the genetically effective population size. *Conservation Biology*, **5**, 325–329.

Santiago E, Caballero A (1998) Effective size and polymorphism of linked neutral loci in populations under directional selection. *Genetics*, **149**, 2105–2117.

Sinnock P (1975) The Wahlund effect for the two-locus model. *American Naturalist*, **109**, 565–570.

Tallmon DA, Luikart G, Waples RS (2004) The alluring simplicity and complex reality of genetic rescue. *Trends in Ecology & Evolution*, **19**, 489–496.

Turner TF, Salter LA, Gold JR (2001) Temporal-method estimates of N_e from highly polymorphic loci. *Conservation Genetics*, **2**, 297–308.

Vitalis R, Couvet D (2001) ESTIM 1.0: a computer program to infer population parameters from one- and two-locus gene identity probabilities. *Molecular Ecology Notes*, **1**, 354–356.

Wang J (2001) A pseudo-likelihood method for estimating effective population size from temporally spaced samples. *Genetic Research*, **78**, 243–257.

Wang J, Ryman N (2001) Genetic effects of multiple generations of supportive breeding. *Conservation Biology*, **15**, 1619–1631.

Wang J, Whitlock MC (2003) Estimating effective population size and migration rates from genetic samples over space and time. *Genetics*, **163**, 429–446.

Waples RS (1989) A generalized approach for estimating effective population size from temporal changes in allele frequency. *Genetics*, **121**, 379–391.

Waples RS (2005) Genetic estimates of contemporary effective population size: to what time periods do the estimates apply? *Molecular Ecology*, **14**, 3335–3352.

Waples RS (2006) A bias correction for estimates of effective population size based on linkage disequilibrium at unlinked gene loci. *Conservation Genetics*, **7**, 167–184.

Waples RS, Do C (1994) Genetic risk associated with supplementation of Pacific salmonids: captive broodstock programs. *Canadian Journal of Fisheries and Aquatic Sciences*, **51** (Suppl. 1), 310–329.

Waples RS, Smouse PE (1990) Gametic disequilibrium analysis as a means of identifying mixtures of salmon populations. *American Fisheries Society Symposium*, **7**, 439–458.

Weir BS (1996) *Genetic Data Analysis II: Methods for Discrete Population Genetic Data*. Sinauer Associates, Sunderland, Massachusetts.

Hitoshi Araki's current research focuses on evolutionary and conservation biology of salmonids and population genetics on plant–pathogen interaction. Robin Waples is interested in developing and applying population genetic principles to real-world problems in ecology, conservation, and management. Mike Blouin's laboratory in Oregon State University focuses on the causes and consequences of genetic structuring, and on applications of methods for parentage analysis in natural populations. The research interests of the authors also include molecular evolution of gene systems, the evolutionary dynamics of genotype–environment interaction, and local adaptation. We are working on a variety of taxa including fish, amphibian, parasites, and plants.
